Lab6

Fitting and Plotting of Binomial distribution & Poisson distribution

BASICS IN PROBABILITY:-

1. If you want to pick five numbers at random from the set 1:50, then you can

```
> sample(1:40,5)
[1] 15 36 10 40 24

> sample(1:40,5)
[1]  3 27  5  8 38

> sample(1:40,5)
[1] 19 17 37 20  6
```

2. Sampling with replacement is suitable for modelling coin tosses or throws of a die.

Roll a die(it gives different results)

```
> sample(1:6,10,replace=TRUE)
[1] 4 4 4 1 4 6 3 5 4 2

> sample(1:6,10,replace=TRUE)
[1] 3 3 4 3 2 1 4 2 6 3
```

3. ## roll 2 die. Even fancier # replace when rolling dice

```
> dice = as.vector(outer(1:6,1:6,paste))
> sample(dice,5,replace=TRUE)
[1] "1 2" "5 2" "4 4" "6 2" "6 4"

> sample(dice,5,replace=TRUE)
[1] "4 3" "3 6" "1 6" "4 2" "4 6"
```

4. #Toss a coin

```
> sample(c("H","T"),10,replace=TRUE)
[1] "H" "H" "H" "H" "T" "T" "T" "H" "H" "H"

> sample(c("H","T"),10,replace=TRUE)
[1] "T" "T" "H" "H" "T" "H" "H" "H" "T" "T"
```

5. # Combination

$\binom{10}{3}$ *OR* $10c_3$

```
> choose(10,3)
[1] 120
```

$\binom{20}{6}$

```
> choose(20,6)
[1] 38760
```

$\binom{30}{5}$

```
> choose(30,5)
[1] 142506
```

6. #permutation (there is no separate permutation function in R)

```
> #P.nk <- factorial(n) / factorial(n-k)
> n=10
> k=5
> P <- factorial(n) / factorial(n-k)
> P
[1] 30240
```

7. Give all binomial coefficients for $\binom{10}{x}$

```
> choose(10,0:10)

 [1]   1  10  45 120 210 252 210 120  45  10   1
```

8. Use a loop to print the first several rows of pasacal's triangle.

```
> for(n in 0:10)print(choose(n,0:n))
[1] 1
[1] 1 1
[1] 1 2 1
[1] 1 3 3 1
[1] 1 4 6 4 1
[1]  1  5 10 10  5  1
[1]  1  6 15 20 15  6  1
[1]  1  7 21 35 35 21  7  1
[1]  1  8 28 56 70 56 28  8  1
 [1]   1   9  36  84 126 126  84  36   9   1
 [1]   1  10  45 120 210 252 210 120  45  10   1
```

Binomial Distribution

The **binomial distribution** is a discrete probability distribution. It describes the outcome of n independent trials in an experiment. Each trial is assumed to have only two outcomes, either success or failure. If the probability of a successful trial is p, then the probability of having x successful outcomes in an experiment of n independent trials is as follows.

$$P[X = x] = \binom{n}{x} p^x q^{n-x}, x=0,1,......n$$

Mean $\mu_1' = np$ **Variance:** $\mu_2 = npq$

Syntax:-

For a binomial(n,p) random variable X, the R functions involve the abbreviation "binom":

```
dbinom(k,n,p)  # binomial(n,p) density at k: Pr(X = k)

pbinom(k,n,p)  # binomial(n,p) CDF at k: Pr(X <= k)

qbinom(P,n,p)  # binomial(n,p) P-th quantile

rbinom(N,n,p)  # N binomial(n,p) random variables

help(Binomial) # documentation on the functions related
               # to the Binomial distribution
```

Problem1. Find the Probability of getting two '4' among ten dice

```
>dbinom(2,size=10,prob=1/6)
[1] 0.29071
```

Problem 2: Find the P(2) by using binomial probability formula

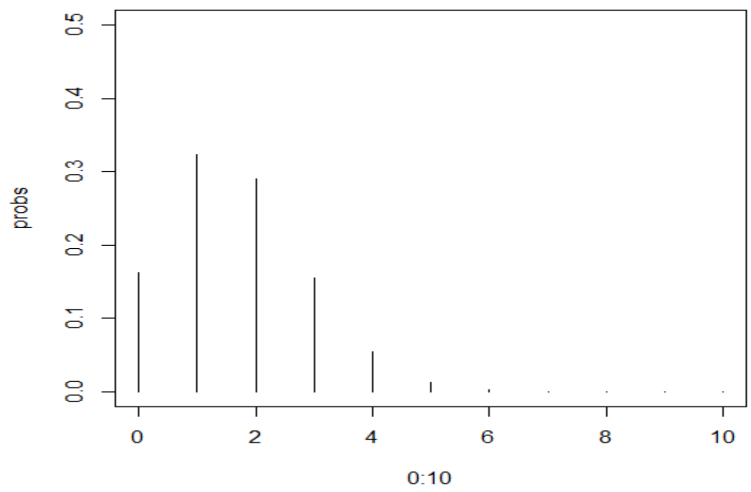
```
> choose(10,2)*(1/6)^2*(5/6)^8
[1] 0.29071
```

Problem 3: Find the table for BIN(n=10,P=1/6)

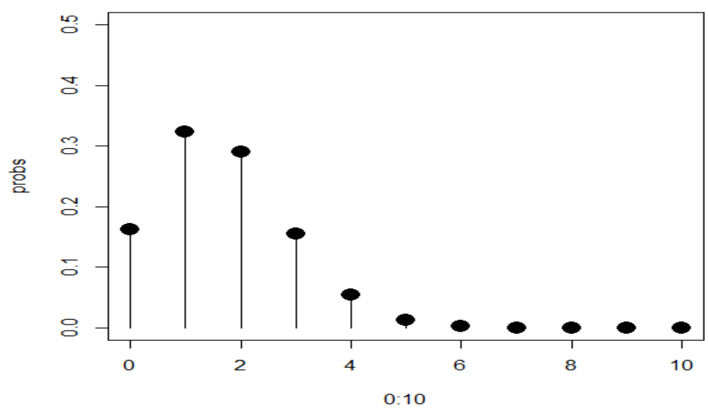
```
> probs=dbinom(0:10,size=10,prob=1/6)
> data.frame(0:10,probs)
   X0.10        probs
1      0 1.615056e-01
2      1 3.230112e-01
3      2 2.907100e-01
4      3 1.550454e-01
5      4 5.426588e-02
6      5 1.302381e-02
7      6 2.170635e-03
8      7 2.480726e-04
9      8 1.860544e-05
10     9 8.269086e-07
11    10 1.653817e-08
> 
```

Problem4: **BINOMIAL PROBABILITY PLOTS :Draw a Plot for the Binomial distribution Bin(n=10,p=1/6)**

>plot(0:10,probs,type="h",xlim=c(0,10),ylim=c(0,0.5))

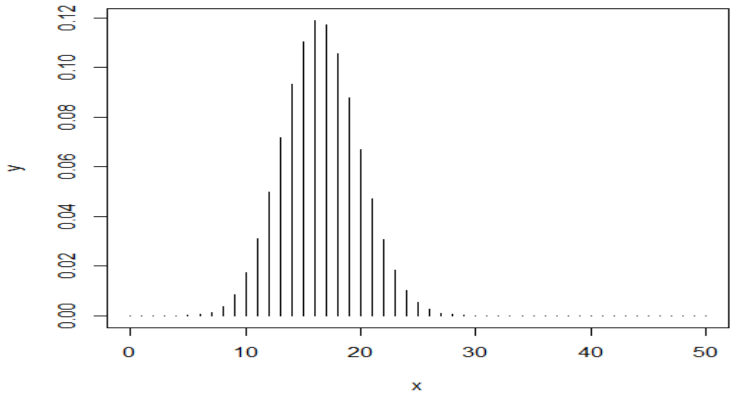


>points(0:10,probs,pch=16,cex=2)



Problem 5: Plot Binomial distribution with n=50 and P=0.33

```
> x=0:50
>(x,size=50,prob=0.33)
> plot(x,y,type="h")
```



Problem 6 : For a Binomial(7,1/4) random variable named X,

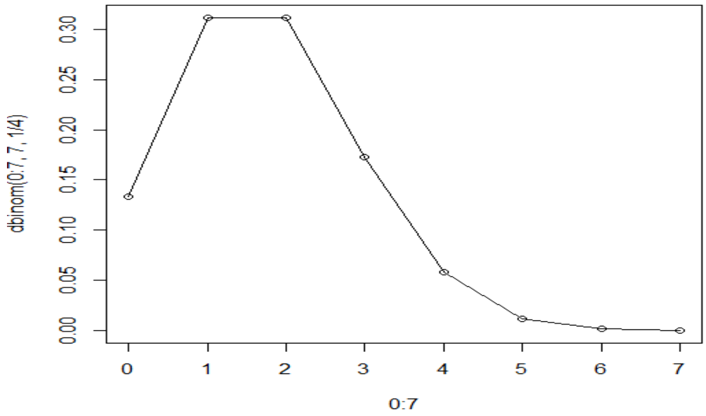
i. Compute the probability of two success
ii. Compute the Probablities for whole space
iii. Display those probabilities in a table
iv. Show the shape of this binomial Distribution

Solution:

```
> dbinom(2,7,1/4)                    # probability of two success
[1] 0.3114624

> dbinom(0:7,7,1/4)                  # probabilities for whole space

[1] 1.334839e-01 3.114624e-01 3.114624e-01 1.730347e-01 5.767822e-02

[6] 1.153564e-02 1.281738e-03 6.103516e-05
```

```
> P=data.frame(0:7,dbinom(0:7,7,1/4))          #probabilities in a Table
> round(P,4)
  X0.7 dbinom.0.7..7..1.4.
1    0              0.1335
2    1              0.3115
3    2              0.3115
4    3              0.1730
5    4              0.0577
6    5              0.0115
7    6              0.0013
8    7              0.0001
```

```
>plot(0:7,dbinom(0:7,7,1/4),type="o")              #shape of the Distribution
```



Problem 7: Suppose there are twelve multiple choice questions in an English class quiz. Each question has five possible answers, and only one of them is correct. Find the probability of having four or less correct answers if a student attempts to answer every question at random.

Solution

Since only one out of five possible answers is correct, the probability of answering a question correctly by random is 1/5=0.2. We can find the probability of having exactly 4 correct answers by random attempts as follows.

R CODE:-

```
> dbinom(4, size=12, prob=0.2)
[1] 0.1329
```

To find the probability of having four or less correct answers by random attempts, we apply the function dbinom with $x = 0,\ldots,4$.

```
> dbinom(0, size=12, prob=0.2) +dbinom(1, size=12, prob=0.2) + dbinom(2, size=12, prob=0
.2) + dbinom(3, size=12, prob=0.2) + dbinom(4, size=12, prob=0.2)
[1] 0.9274
```

Or

```
> sum(dbinom(x=0:4,size=12,prob=0.2))

[1] 0.9274445
```

or

Alternatively, we can use the cumulative probability function for binomial distribution pbinom.

```
> pbinom(4, size=12, prob=0.2)
[1] 0.92744
```

The probability of four or less questions answered correctly by random in a twelve question multiple choice quiz is 92.7%.

Problem 8: If 10% of the Screws produced by an automatic machine are defective, find the probability that out of 20 screws selected at random, there are

(i) Exactly 2 defective
(ii) At least 2 defectives
(iii) Between 1 and 3 defectives (inclusive)

Code:-

(i)
```
> # Exactly 2 defective P=0.10 n=20
> dbinom(2,20,0.10)
[1] 0.2851798
```
(ii)
```
> 1-dbinom(1,20,0.10)
[1] 0.7298297
```
(iii)
```
> #Between 1 and 3 defectives (inclusive)
> x=sum(dbinom(1:3,20,0.10))
> x
[1] 0.74547
```

Relationship between mean and variance :-

Problem 9: Show that Binomial distribution variance is less than mean with Binomial variable follows (7,1/4)

```
> n=7
> p=1/4
> x=dbinom(0:7,n,p)
> x
[1] 1.334839e-01 3.114624e-01 3.114624e-01 1.730347e-01 5.767822e-02
[6] 1.153564e-02 1.281738e-03 6.103516e-05
> Ex=sum(x*p)
> Ex
[1] 0.25
> var=sum((x-Ex)^2*x)
> var
[1] 0.008062817
```

THE POISON DISTRIBUTION:

If the number of Bernoulli trials of a random experiment is fairly large and the probability of success is small it becomes increasingly difficult to compute the binomial probabilities. For values of n and p such that $n \geq 150$ and $p \leq 0.05$, the poisson distribution serves as an excellent approximation to the binominal distribution.

The random variable X is said to follow the Poisson distribution if and only if

$$p[X = x] = \frac{e^{-\lambda}\lambda^{x}}{\lfloor x}, x = 0,1,2,.....$$

Assumptions:-

1. Number of Bernoulli trials (n) is indefinitely large, $(n \to \infty)$
2. The trials are independent.
3. Probability of success (p) is very small, $(p \to 0)$

   λ = np is constant, $\lambda = np \Rightarrow p = \frac{\lambda}{n}$
4. Mean and variance in poison distribution are equal

Syntax:-

```
dpois(x, lambda, log = FALSE)
ppois(q, lambda, lower.tail = TRUE, log.p = FALSE)
qpois(p, lambda, lower.tail = TRUE, log.p = FALSE)
rpois(n, lambda)
```

Problem 1:

a. #P(x=5) with parameter 7
```
> dpois(x=5,lambda=7)
[1] 0.1277167
```

b. #P(x=0)+P(x=1)+.........+P(x=5)
```
> dpois(x=0:5,lambda=7)

[1] 0.000911882 0.006383174 0.022341108 0.052129252 0.091226192 0.127716668
```

c.
```
> #P(x<=5)
> sum(dpois(0:5,lambda=7))
[1] 0.3007083
```
Or
```
> ppois(q=4,lambda=7,lower.tail=T)
[1] 0.1729916
```

d.
```
> ppois(q=12,lambda=7,lower.tail=F)

[1] 0.02699977
```

Problem 2 : Check the relationship between mean and variance in Poisson distribution(4) with n=100

```
> X.val=0:100
> P.val=dpois(X.val,4)
> EX=sum(X.val*P.val)            #mean
> EX
[1] 4
> sum((X.val-EX)^2*P.val)        #variance
[1] 4
```

Problem 3 : Compute Probabilities and cumulative probabilities of the values between 0 and 10 for the parameter 2 in poisson distribution.

```
> dpois(0:10,2)                          # probabilities

 [1] 1.353353e-01 2.706706e-01 2.706706e-01 1.804470e-01 9.022352e-02
 [6] 3.608941e-02 1.202980e-02 3.437087e-03 8.592716e-04 1.909493e-04
[11] 3.818985e-05
```

Or

```
> P=data.frame (0:10,dpois(0:10,2))
> round (P,4)
```

	X0.10	dpois.0.10..2.
1	0	0.1353
2	1	0.2707
3	2	0.2707
4	3	0.1804
5	4	0.0902
6	5	0.0361
7	6	0.0120
8	7	0.0034
9	8	0.0009
10	9	0.0002
11	10	0.0000

```
> ppois(0:10,2)                          # cumulative probabilities

 [1] 0.1353353 0.4060058 0.6766764 0.8571235 0.9473470 0.9834364 0.9954662
 [8] 0.9989033 0.9997626 0.9999535 0.9999917
```

Or

```
> P=data.frame(0:10,ppois(0:10,2))
> round(P,4)
   X0.10 ppois.0.10..2.
1      0         0.1353
2      1         0.4060
3      2         0.6767
4      3         0.8571
5      4         0.9473
6      5         0.9834
7      6         0.9955
8      7         0.9989
9      8         0.9998
10     9         1.0000
11    10         1.0000
```

Problem 3: Poisson distribution with parameter '2'

1. How to obtain a sequence from 0 to 10
2. Calculate P(0),P(1),...,P(10) when lambda =2 and Make the output prettier
3. Find P(x <= 6)
4. Sum all probabilities
5. Find P(Y>6)
6. Make a table of the first 11 Poisson probs and cumulative probs when # mu=2 and make the output prettier
7. Plot the probabilities Put some labels onthe axes and give the plot a title:

a. > 0:10 #sequence from 0:10

```
[1]  0  1  2  3  4  5  6  7  8  9 10
```

b. > round(dpois(0:10, 2), 3)

```
[1] 0.135 0.271 0.271 0.180 0.090 0.036 0.012 0.003 0.001 0.000 0.000
```

c. > ppois(6, 2) # Find P(x <= 6)

```
[1] 0.9954662
```

d. > sum(dpois(0:6, 2)) # Sum all probabilities

```
[1] 0.9954662
```

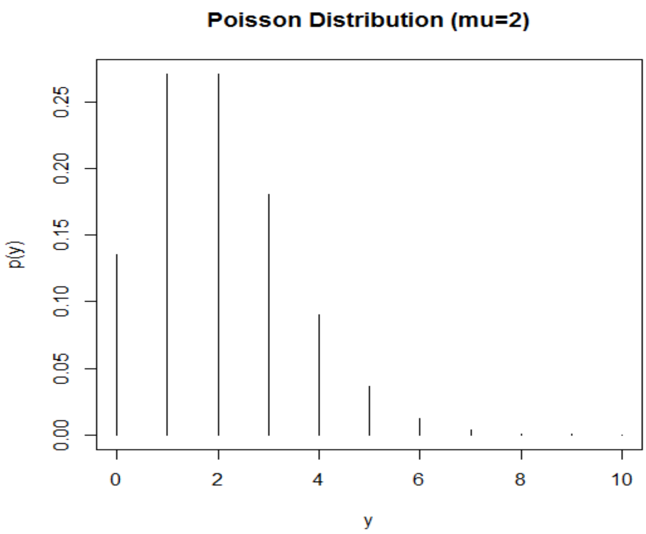
e. > 1 - ppois(6, 2) # Find P(Y>6)

```
[1] 0.004533806
```

f.

```
> round(cbind(0:10, dpois(0:10,2), ppois(0:10,2)), 3)
      [,1]  [,2]  [,3]
 [1,]    0 0.135 0.135
 [2,]    1 0.271 0.406
 [3,]    2 0.271 0.677
 [4,]    3 0.180 0.857
 [5,]    4 0.090 0.947
 [6,]    5 0.036 0.983
 [7,]    6 0.012 0.995
 [8,]    7 0.003 0.999
 [9,]    8 0.001 1.000
[10,]    9 0.000 1.000
[11,]   10 0.000 1.000
```

g. plot(0:10,dpois(0:10,2),type="h",xlab="y",ylab="p(y)",main="Poisson Distribution (mu=2)")



Practice Problems:-(Binomial distribution)

1. For a random variable X with a binomial(20,1/2) distribution, find the following probabilities.
   - (i). Find Pr(X < 8)
   - (ii). Find Pr(X > 12)
   - (iii) Find Pr(8 <= X <= 12)

2. For a binomial(200,1/2) distribution:
   - (i) Find Pr(X < 80)
   - (ii) Find Pr(X > 120)
   - (iii) Find Pr(80 <= X <= 120)

3. For a binomial(2000,1/2) distribution:
   Find Pr(X < 800)
   Find Pr(X > 1200)
   Find Pr(800 <= X <= 1200)

4. Let X be the number of heads in 10 tosses of a fair coin.
   1. Find the probability of getting at least 5 heads (that is, 5 or more).
   2. Find the probability of getting exactly 5 heads.
   3. Find the probability of getting between 4 and 6 heads, inclusive

5. Suppose our random variable X is Poisson with lambda = 12.33.

   1. What is the probability of 15 or fewer occurrences? P(X <= 15)
   2. What is the probability of EXACTLY 6 occurrences? P(X = 6)
   3. What is the probability of more than 15 occurrences? P(X > 15)
   4. What is the probability of 15 or more occurrences? P(X >= 15)
   5. What is the probability of 8, 9, or 10 occurrences? P(8 <= X <= 10)

Compare binomial distribution and Poisson distribution

6. Let X be the number of heads in 100 tosses of a fair coin.
7. Let X be the number of heads in 1000 tosses of a fair coin.

Challenging Experiments:

Problem 1: A recent national study showed that approximately 55.8% of college students have used Google as a source in at least one of their term papers. Let X equal the number of students in a random sample of size n = 42 who have used Google as a source:

1. How is X distributed?
2. Sketch the probability mass function (roughly).
3. Sketch the cumulative distribution function (roughly).
4. Find the probability that X is equal to 17.
5. Find the probability that X is at most 13.
6. Find the probability that X is bigger than 11.
7. Find the probability that X is at least 15.
8. Find the probability that X is between 16 and 19, inclusive
9. Give the mean of X, denoted IE X.
10. Give the variance of X.

11. Give the standard deviation of X.
12. Find IE(4X + 51:324)
13. Compare mean and variance

Problem 2: (Traffic accident problem)

The number of traffic accidents that occur on a particular stretch of road during a month follows a Poisson distribution with a mean of 7.6.

1. Find the probability that less than three accidents will occur next month on this stretch of road.
2. Find the probability of observing exactly three accidents on this stretch of road next month.
3. Find the probability that the next two months will both result in four accidents each occurring on this stretch of road.
4. Check the mean and variance of the poisson distribution
5. Plot the Poisson distribution and compare with binomial distribution